

# Voice gender differences and separation of simultaneous talkers in cochlear implant users with residual hearing

Anisa S. Visram, Karolina Kluk, and Colette M. McKay

*School of Psychological Sciences, University of Manchester,  
Manchester M13 9PL, United Kingdom*

*karolina.kluk@manchester.ac.uk, colette.mckay@manchester.ac.uk,  
anisa.visram@manchester.ac.uk*

**Abstract:** Perception of a target voice in the presence of a competing talker, of same or different gender as the target, was investigated in cochlear implant users, in implant-alone and bimodal (acoustic hearing in the non-implanted ear) conditions. Recordings of two male and two female talkers acted as targets and maskers, to investigate whether bimodal benefit increased for different compared to same gender target/maskers due to increased ability to perceive and utilize fundamental frequency and spectral-shape differences. In both listening conditions participants showed benefit of target/masker gender difference. There was an overall bimodal benefit, which was independent of target/masker gender difference.

© 2012 Acoustical Society of America

PACS numbers: 43.66.Ts, 43.71.Ky [QJF]

Date Received: May 18, 2012 Date Accepted: June 25, 2012

## 1. Introduction

Cochlear implant (CI) users with residual acoustic hearing can achieve significant speech understanding benefit by combining acoustic and electric hearing, and this bimodal benefit has often been attributed to the perception of fundamental frequency (F0) in the acoustic signal (Brown and Bacon, 2009; Cullington and Zeng, 2010; Zhang *et al.*, 2010), though broader spectral cues such as first formant information may also be important (Kong and Carlyon, 2007; Sheffield and Zeng, 2012; Visram *et al.*, 2012). One of the benefits of the F0 coding provided by the acoustic signal may be an increased ability to segregate a talker from background sounds. Previous research suggests that some CI users have difficulty in making use of gender and F0 differences to aid separation of simultaneous voices (Stickney *et al.*, 2004; Stickney *et al.*, 2007) while others may get similar benefits to normally hearing controls (Cullington and Zeng, 2008). Typically, those CI users who have some useable residual hearing will have better thresholds at low frequencies than at high frequencies, and thus are likely to have access to F0 information acoustically, at least when aided. Qin and Oxenham (2006) investigated simultaneous vowel perception as a function of F0 separation for normally-hearing listeners listening to a simulation of combined CI and acoustic hearing. Vowels were artificially synthesized to have different F0s, while keeping the spectral shape the same for each vowel. They found that when listening with the CI simulation alone, there was no effect of F0 separation between the vowels. However, when low-frequency acoustic hearing was added to the simulation there was a benefit for increasing F0 separation, similar to that seen in the unprocessed condition, suggesting that improved F0 representation helped to segregate the vowels. Testing CI users with residual hearing, Kong *et al.* (2005) found a benefit of bimodal listening over CI-alone for perception of male-talker target sentences in the presence of a competing talker. The amount of bimodal benefit was greater for a female than for a

male masker, suggesting that the F0 and/or spectral differences between targets and maskers of different genders helped segregation in the bimodal condition. [Psychny \*et al.\* \(2011\)](#) investigated bimodal speech perception with competing talkers, by artificially manipulating differences in F0 and vocal tract length (i.e., relative location of formant peaks) between target and masker. They found that F0 differences, but not vocal tract length differences, helped to segregate target and masker in both CI-alone and bimodal conditions. There was an overall bimodal benefit, but this benefit was independent of segregation ability. [Cullington and Zeng \(2011\)](#) found that bilateral and bimodal cochlear implant users performed similarly on speech perception with a competing talker, and that both groups perceived the male target voice better when the masker was female than when the masker was also male.

Given conflicting results from previous studies, the present study investigated whether implant users can use gender-related voice differences to help segregate competing talkers and whether residual acoustic hearing can help them to do so. The study used both male and female target and masker talkers to limit the confounding effects of audibility within the range of residual hearing. Also, each voice acted as both target and masker in separate runs, to minimize the baseline effects of voice intelligibility and masking effectiveness of individual voices.

## 2. Methods

### 2.1 Participants

Nine experienced CI users with low-frequency residual hearing in the non-implanted ear took part. Audiometric thresholds in the non-implanted ear are shown in Fig. 1, and demographic data for the participants are given in Table 1. All participants gave signed consent to take part in the study, which was ethically approved by the UK National Research Ethics Service (NRES).

### 2.2 Stimuli and stimulus presentation

Test materials were two male (M1, M2) and two female (F1, F2) recordings of the coordinate response measure (CRM) sentences, of the structure “ready <<call sign>> go to <<color>> <<number>> now” ([Moore, 1981](#)). There were eight different options for call sign, four options for color, and eight options for number. The average F0s of the sentences were as follows: M1: 119 Hz, M2: 104 Hz, F1: 241 Hz, F2: 209 Hz. There were four different male/female, target/masker combinations: MM: male target with male

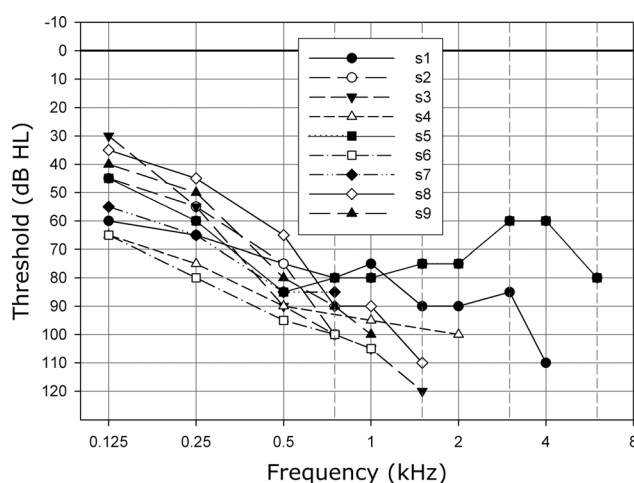


Fig. 1. Participants' audiometric thresholds in the non-implanted ear.

Table 1. Demographic data: Age in years; duration of implant use in years and months at start of testing; use of contralateral hearing aid in daily life. Y: yes (regular use), N: no use, O: occasional use.

ID	Age	Implant use	Hearing aid use	Processor	Processing strategy	Etiology
S1	69	5y9mo	Y	Esprit 3G	ACE	Familial progressive
S2	79	6y8mo	O	Harmony	HiRes-P w/Fidelity	Idiopathic progressive
S3	76	3y8mo	N	Freedom	ACE	Familial progressive
S4	47	2y10mo	N	Freedom	ACE	Head injury
S5	78	10mo	Y	CP810	ACE	Idiopathic progressive
S6	50	5y10mo	N	Freedom	ACE	Ototoxicity
S7	72	6mo	Y	Opus 2	FSP	Familial progressive
S8	57	7mo	Y	Opus 2	FSP	Idiopathic progressive
S9	64	3y6mo	N	Opus 2	FSP	Idiopathic progressive

masker; MF: male target with female masker; FF: female target with female masker; FM: female target with male masker. Additionally, participants were tested in two different target-masker groupings: TM1 in which M1 and F1 were targets with M2 and F2 as maskers; and TM2 in which target and maskers were reversed.

Sentences were played to the implant via direct audio input and to the non-implanted ear via an E-A-R 3A insert earphone. Acoustic stimuli were initially passed through a fast Fourier transform filter to mimic the participant's NAL-RP prescription, then adjusted to the subjective most comfortable level, using output compression if necessary to avoid peak clipping. CI stimuli were also adjusted to the subjective most comfortable level, and if necessary level adjustments were made to achieve a comfortable level when listening bimodally. For subject s2 only, stimuli were played to the implant through a supra-aural headphone placed over the t-mic of the speech processor (a microphone designed to sit at the entrance of the ear canal). This was done as the experimenters were advised the speech processor should be directly connected to handheld battery-powered devices only.

### 2.3 Procedures

Subjects were asked to follow the target talker who named the call sign "Baron," and to identify the color and number said by that talker. The target CRM sentence was always presented with a masker CRM sentence that had a call sign other than "Baron." A one-up one-down adaptive procedure was used to find the target-to-masker ratio (TMR) at which participants correctly identified both target keywords 50% of the time. The starting TMR was +20 dB, changing in 8 dB steps until the first turning point, 4 dB steps until the third turning point, and 2 dB steps for the next 12 turning points. While the TMR was above 0 dB the target was fixed and the masker adjusted, but while the TMR was below 0 dB the target was adjusted keeping the masker level fixed. This was done to limit the target and masker input levels to a range that avoided distortion or clipping. The TMR<sub>50%</sub> was the mean TMR at the last 12 turning points. Tracks that did not appear to asymptote or with large variability (a standard deviation of the mean of upward turning points of over 3 dB) were rejected and repeated. If time allowed, tracks with a standard deviation of upward turning points of over 2 dB were repeated and an average of both runs was taken as the final score. Participants were tested over four sessions in total. In each session, a single target-masker grouping (TM1 or TM2) and a single run of every target/masker condition within that grouping (MM, MF, FF, FM) were tested, in the two different listening modes (CI alone and bimodal). Each target-masker grouping was thus tested twice, the order of testing within the group being reversed for the repeat session. In the later sessions, if time allowed, any conditions with large discrepancies between the TMR<sub>50%</sub> for repeat runs

across sessions (of 3 dB or more) were repeated once more, and the average of all three runs used for that score. The order of testing was counterbalanced as far as possible within and between participants. Results for the male/female, target/masker conditions were averaged across repeat sessions and across the two target-masker groupings.

### 3. Results

Figure 2 shows the individual and mean scores across target/masker conditions for subjects in the CI-alone and bimodal conditions. A two-way repeated-measures analysis of variance was performed, with listening mode and target/masker condition as factors. There was a significant effect of listening mode [ $F(1, 8) = 26.461$ ,  $p < 0.001$ ] showing an overall bimodal benefit compared to CI-alone. There was also a significant effect of target/masker condition [ $F(3, 8) = 15.537$ ,  $p < 0.001$ ]. *Post hoc* comparisons using the Holm-Sidak method showed that significant differences [ $p < 0.05$ ] were present between all same-gender and mixed-gender conditions (i.e., MM vs MF and FM; FF vs MF and FM) with an advantage to mixed-gender conditions, but showed no differences between same-gender conditions (i.e., MM vs FF) or mixed-gender conditions (i.e., FM vs MF). There was no significant interaction between listening mode and target/masker condition ( $p = 0.584$ ) implying that the bimodal benefit seen in this task was independent of target/masker condition. In other words, the advantage of the mixed over fixed gender conditions was present equally in both CI-alone and bimodal conditions.

### 4. Discussion

In contrast to the results of [Stickney \*et al.\* \(2004\)](#) but in agreement with those of [Cullington and Zeng \(2008, 2010\)](#), participants were able to use gender differences between target and masker in the CI-alone condition. This is also in agreement with results of [Psychny \*et al.\* \(2011\)](#) who showed that CI users could use artificial increases in F0 in the masker compared to target sentence to obtain masking release, and contrasts with [Stickney \*et al.\* \(2007\)](#) who showed benefit of artificial F0 differences for normally-hearing subjects but not CI users. The design of the study was such that speech scores were averaged over two different target-masker groupings, allowing each voice to act both as target and masker in separate runs. We can therefore discount the effect of individual talker intelligibility differences in our results and have confidence that these listeners were able to use gender voice differences in both CI alone and bimodal conditions to help separate target from masker. A possible reason for the differences between our results and those of [Stickney \*et al.\* \(2004\)](#) and [Stickney \*et al.\* \(2007\)](#) is the range of TMRs at which the test was performed, with relatively lower TMRs in our study. The average TMR<sub>50%</sub> obtained in the current experiment was close to 0 dB (compared to an average TMR of 10 dB for [Stickney \*et al.\*](#)), at which we could expect a large amount of informational masking, and a need to use voice difference cues to distinguish the two voices. Furthermore, in their studies, the influence of informational masking may have been somewhat reduced by the same masker sentence being used in each trial.

The amount of bimodal benefit was *not* related to target/masker condition, suggesting the use of acoustic hearing did not increase the ability of CI users to use F0 or spectral differences to perceptually separate competing talkers with differing genders. One explanation for the lack of extra bimodal benefit seen in different-gender conditions compared to same-gender ones, could be the difficulty in using F0 to group low-frequency components from the acoustic signal with high-frequency components from the implant signal. Congruent F0 information between high- and low-frequency portions of the target may be important for such grouping in normal hearing ([Bird and Darwin, 1998](#)) but it does not seem to be crucial for simulated bimodal speech benefit ([Brown \*et al.\*, 2010](#); [Kong and Carlyon, 2007](#)). This could suggest the mechanism in normal hearing by which different frequency bands are streamed by use of F0

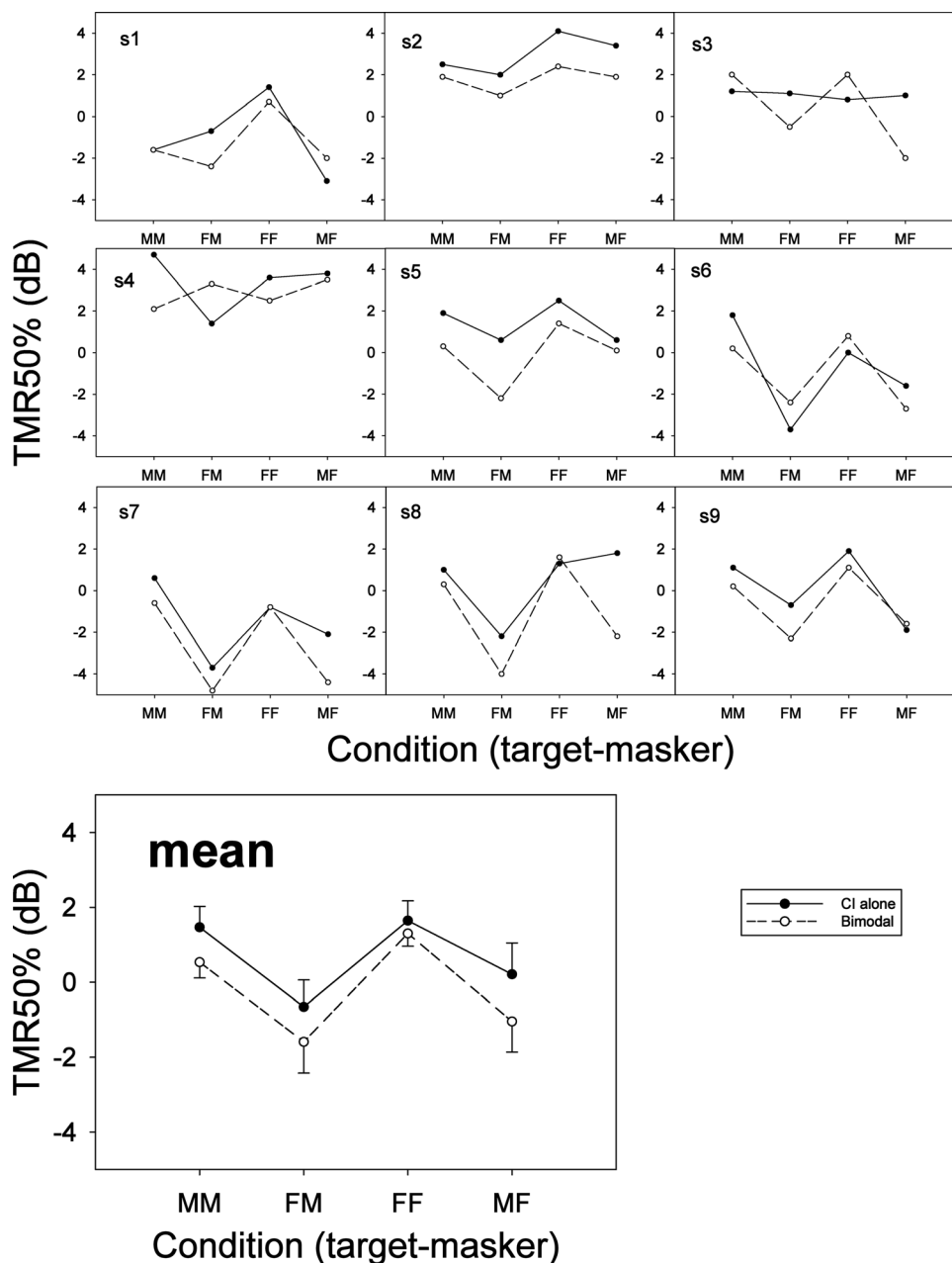


Fig. 2. Individual and mean TMR50% scores for all subjects across target/masker conditions, for CI alone (filled circles) and bimodal (open circles) listening modes. Results for target-masker groups TM1 and TM2 are averaged together. Error bars show the standard error of the mean.

information, is not effective for bimodal listeners. A further explanation is that these CI users were able to extract adequate F0 information from their CI alone to make use of the target-masker F0 separation for segregation of voices. A paradigm which tests a range of target-masker F0 differences may better highlight any differences in segregation based on F0 perception between CI alone and bimodal conditions. Additionally, it is plausible that for large F0 separations, such as those between the male and female talkers, the F0s of individual voices may fall into separate CI channels.

This may assist CI listeners, particularly in tracking the lower-pitched voice, as the lower channel would contain information from that voice only.

Acoustic perception of F0 may be useful for bimodal listeners in other ways, rather than increasing segregation ability. For example, F0 provides a cue to voicing and may provide useful information about the energy of the target and maskers signals to aid listening in the dips of the masker (Brown and Bacon, 2009; Kong and Carlyon, 2007; Sheffield and Zeng, 2012). Furthermore wider spectral cues such as F1 information may be useful for bimodal speech benefit (Kong and Carlyon, 2007; Sheffield and Zeng, 2012; Visram *et al.*, 2012).

## 5. Conclusions

Participants performed better in mixed-gender than in same-gender target/masker conditions. This effect is consistent with an advantage for mixed-gender target/maskers rather than individual voice intelligibility, as scores were averaged over conditions in which the different voices acted as both targets and maskers. Bimodal listening significantly improved perception of the target speech in the presence of a competing voice, with a 1 dB TMR mean benefit. The amount of bimodal benefit was not related to target/masker condition. Therefore the bimodal benefit seen, for these subjects and listening conditions, was not due to better ability to separate talkers based on voice-gender differences with the help of the acoustic signal.

## Acknowledgments

This work was financially supported by the University of Manchester. We wish to thank all the participants for their time and patience, and three anonymous reviewers for helpful comments on an earlier version of the manuscript.

## References and links

- Bird, J., and Darwin, C. J. (1998). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.
- Brown, C. A., and Bacon, S. P. (2009). "Achieving electric-acoustic benefit with a modulated tone," *Ear Hear.* **30**, 489–493.
- Brown, C. A., Scherrer, N. M., and Bacon, S. P. (2010). "Shifting fundamental frequency in simulated electric-acoustic listening," *J. Acoust. Soc. Am.* **128**, 1272–1279.
- Cullington, H. E., and Zeng, F. G. (2008). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," *J. Acoust. Soc. Am.* **123**, 450–461.
- Cullington, H. E., and Zeng, F. G. (2010). "Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear," *Ear Hear.* **31**, 70–73.
- Cullington, H. E., and Zeng, F. G. (2011). "Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification," *Ear Hear.* **32**, 16–30.
- Kong, Y. Y., and Carlyon, R. P. (2007). "Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation," *J. Acoust. Soc. Am.* **121**, 3717–3727.
- Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (2005). "Speech and melody recognition in binaurally combined acoustic and electric hearing," *J. Acoust. Soc. Am.* **117**, 1351–1361.
- Moore, T. (1981). "Voice communication jamming research," in *AGARD Conference Proceedings 331: Aural Communication in Aviation*, Neuilly-Sur-Seine, France, pp. 2:1–2:6.
- Psychny, V., Landwehr, M., Hahn, M., Walger, M., von Wedel, H., and Meister, H. (2011). "Bimodal hearing and speech perception with a competing talker," *J. Speech Lang. Hear. Res.* **54**, 1400–1415.
- Qin, M. K., and Oxenham, A. J. (2006). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," *J. Acoust. Soc. Am.* **119**, 2417–2426.
- Sheffield, B. M., and Zeng, F. G. (2012). "The relative phonetic contributions of a cochlear implant and residual acoustic hearing to bimodal speech perception," *J. Acoust. Soc. Am.* **131**, 518–530.

- Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G. (2007). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," *J. Acoust. Soc. Am.* **122**, 1069–1078.
- Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Visram, A. S., Azadpour, M., Kluk, K., and McKay, C. M. (2012). "Beneficial acoustic speech cues for cochlear implant users with residual acoustic hearing," *J. Acoust. Soc. Am.* **131**, 4042–4050.
- Zhang, T., Dorman, M. F., and Spahr, A. J. (2010). "Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation," *Ear Hear.* **31**, 63–69.